



# Considérations épistémologiques sur la modélisation mathématique en biologie

Philippe Huneman, Sébastien Dutreuil

## ► To cite this version:

Philippe Huneman, Sébastien Dutreuil. Considérations épistémologiques sur la modélisation mathématique en biologie. Thierry Hoquet et Francesca Merlin. Précis de Philosophie de la biologie, , 2014, 978-2-311-40019-9. hal-01109182

**HAL Id: hal-01109182**

**<https://hal.science/hal-01109182>**

Submitted on 31 Jan 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Considérations épistémologiques sur la modélisation mathématique en biologie.

Philippe Huneman (IHPST, Paris).

Sébastien Dutreuil, (Université Paris 1 Panthéon-Sorbonne, UMR8590, IHPST, Paris)

Paru dans *Précis de philosophie de la biologie*, ed: Thierry Hoquet et Francesca Merlin, Vuibert, 2014.

### Introduction.

Pendant longtemps, le biologiste a considéré les mathématiques avec circonspection. Les phénomènes du vivant semblaient trop flexibles, plastiques, variables ou individuels pour se plier à la rigidité des cadres mathématiques. Le caractère quantitatif censément propre à la méthode mathématique paraissait inapte à saisir le propre du vivant, qui relèverait d'une réflexion plus qualitative. L'histoire naturelle, de Théophraste aux grands naturalistes comme Linné, Buffon ou Jussieu, inventoriait le vivant à partir de descriptions précises des différentes espèces et variétés ; la physiologie à partir de Haller puis Bichat entre 1750 et 1800 a introduit l'usage systématique de l'expérimentation et donc certains protocoles quantitatifs, mais ce n'est qu'à partir du 20<sup>ème</sup> siècle que l'usage des mathématiques est devenu une évidence pour la biologie.

Cette résistance aux mathématiques se fondait parfois sur une inspiration vitaliste ou antimécaniste, comme c'était le cas pour certains physiologistes du 18<sup>ème</sup> siècle tels que Bichat ou les vitalistes montpelliérains, mais il s'agit d'une idée beaucoup plus générale, souvent partagée par ceux-là mêmes qui s'élèvent contre le vitalisme, l'organicisme, etc. C'est ainsi que Claude Bernard, sévère critique de tout vitalisme et partisan d'une méthode scientifique identique pour la physique et la biologie protestait encore en 1859 contre l'usage des statistiques en physiologie<sup>1</sup>.

Les philosophes ont élaboré de nombreuses raisons pour soutenir cette méfiance : pour Aristote les mathématiques sont une pure abstraction inapte à comprendre le mouvement (*kinésis*), lequel définit la nature en son entier et s'appuie toujours sur une cause finale, une caractéristique dont les organismes sont des exemples paradigmatiques ; pour Hegel dans la *Science de la logique* la vie réalise une forme d'être incorporant la contradiction (être le même tout en devenant autre), de sorte que les mathématiques, gouvernées par le principe de non-contradiction, sont trop sommaires pour en rendre compte (Hegel 1816). Plus récemment, Bergson est célèbre pour avoir affirmé que les mathématiques,

---

<sup>1</sup> Entre autres citations de l'*Introduction à l'étude de la médecine expérimentale*, dans le chapitre II section 9: "Une autre forme d'application très fréquente des mathématiques à la biologie se trouve dans l'usage des moyennes ou dans l'emploi de la statistique qui, en médecine et en physiologie, conduisent pour ainsi dire nécessairement à l'erreur. Il y a sans doute plusieurs raisons pour cela ; mais le plus grand écueil de l'application du calcul aux phénomènes physiologiques, est toujours au fond leur trop grande complexité qui les empêche d'être définis et suffisamment comparables entre eux." (Bernard 1865, p.181)

forme analytique propre à l'« intelligence » et incapable de saisir la « durée » authentique, ne peuvent comprendre le propre du vivant, l'élan vital, qui requiert l'usage de ce qu'il appelle « intuition » (cf. Bergson 1907).

Néanmoins, contre la vision dominante il a existé, depuis le début du 20<sup>ème</sup> siècle un courant minoritaire de biologistes pour lesquels les mathématiques constituent une voie d'accès à la compréhension des phénomènes du vivant. Parmi les plus connus, d'Arcy Thompson dans *On Growth and form* affirme que les formes vivantes et leur distribution obéissent à des lois mathématiques parfois très simples : ainsi, le rapport entre la différence entre la croissance de la surface ( $x^2$ ) et du volume ( $x^3$ ) d'un corps rend compte des propriétés différentes des gros et des petits organismes : les fourmis peuvent tomber de haut sans se tuer, pas les chevaux. Ou encore, la phyllotaxie des plantes suit une séquence définie par la suite de Fibonacci. Un peu plus tard, Nicolas Rashevsky (1938) développe des principes de « biophysique mathématique », discipline analogue à la physique mathématique mais portant sur les phénomènes biologiques : ainsi la cellule, analogue à une gouttelette soumise à des forces de tension, obéit à des équations qui permettent de prédire le point et le schéma de la division cellulaire. Dans les années 1970 Robert Rosen proposera une théorie mathématique de la nature de l'organisme, tentant de modéliser la causalité circulaire qui caractérise ce type d'entités (Rosen 1991) et qui avait été régulièrement soulignée par des philosophes tels que Kant ou des biologistes comme Cuvier ou Bichat: ici encore, on fait le pari de l'aptitude de la formalisation mathématique à constituer la théorie adéquate à la compréhension du vivant comme tel.

D'Arcy Thomson, Rashevsky ou Rosen représentent des positions extrêmes au sens où la biologie devient elle-même intégralement une théorie mathématique. Le rapport des mathématiques à la biologie à partir du 20<sup>ème</sup> siècle n'est toutefois pas unanimement caractérisable par cette position, pour laquelle les mathématiques seraient l'expression nécessaire des vérités biologiques, et non des outils spécifiques pour résoudre des problèmes biologiques précis. Cette seconde vision, moins forte en termes d'engagements réalistes quant à la valeur ontologique des mathématiques en biologie, est sans doute celle que partagerait la majorité des biologistes. Elle correspond en gros à l'idée que les biologistes font des modèles des phénomènes et que ceux-ci s'expriment en termes mathématiques, même si l'idée que toute la biologie puisse relever d'un ensemble de lois mathématiquement exprimables n'est pas acceptée par de nombreux biologistes.

Le présent chapitre commencera par des généralités sur les modèles mathématiques. Il distinguera les modèles dits « phénoménologiques » ou modèles de patterns, et les modèles dits « mécanistes » ou modèles de processus, et soulignera des réquisits qui caractérisent chacun d'eux. La seconde section examinera caractères et problèmes communs aux modèles de patterns, et la suivante se tournera vers les modèles de processus. Suivant la distinction utile faite par Ernst Mayr (1961) entre *biologie évolutive*, qui étudie les causes ultimes (situées dans le passé d'une espèce), et *biologie fonctionnelle*, qui étudie les causes prochaines, situées dans la durée de vie d'un organisme individuel, on commencera par considérer certaines spécificités de modèles mathématiques

dans la biologie évolutive, puis on se penchera sur des exemples de modèles en biologie fonctionnelle. Nous soulignerons dans chaque cas les spécificités des mathématiques utilisées ; une dernière section interrogera la place des mathématiques dans les diverses explications biologiques. Le chapitre insistera de manière générale sur les raisons pour laquelle le recours aux mathématiques est devenu inévitable dans la modélisation biologique.

### **Modèles : patterns, processus, mathématiques.**

La philosophie des sciences dès sa naissance institutionnelle à la fin du 19<sup>ème</sup> siècle s'est interrogée sur le propre de *l'explication* scientifique : aux controverses entre Whewell et Mill sur le caractère inductif ou non de l'explication scientifique, succèdent les interrogations sur le modèle nomologique-déductif d'Hempel, et une littérature abondante sur le caractère causal, unificateur, etc., de l'explication scientifique (voir Salmon 1989 pour un panorama sur ce point). La dernière décennie a vu une inflation des réflexions sur la notion même de *modèle* scientifique, dont le caractère central dans la pratique scientifique elle-même est incontestable (e.g Giere, 1988, Morgan et Morrison, 1999) – les raisons de cette attention particulière sur le modèle sont nombreuses, parfois suscitées par une mise entre parenthèses de la question du réalisme scientifique, ou bien par l'émergence de la conception dite « sémantique » des théories (e.g Van Fraassen, 1980, voir Vorms 2011 pour une présentation). Ainsi, quelle que soit la vision que l'on adopte de l'explication scientifique ou de la théorie scientifique (autre objet majeur du questionnement des premiers philosophes des sciences), on reconnaîtra que théories comme explications scientifiques ont recours à des modèles ; or ces modèles sont assez souvent à caractère mathématique.

Les scientifiques distinguent souvent modèles « phénoménologiques », et modèles « mécanistes » ; en biologie, en particulier en biologie évolutive, on parle plutôt de « modèle de pattern » et de « modèle de processus ». Le processus produit le pattern, tout comme le modèle mécaniste vise à saisir la raison du phénomène représenté par le modèle phénoménologique ; les deux couples de concepts ne correspondent pas exactement mais pour notre propos on peut les prendre comme équivalents.

Si l'on considère que des données sont des échantillons tirés d'une réalité – des enregistrements de position de flamants roses bagués, des registres d'hôpitaux concernant les patients infectés par un virus, des relevés de température sur plusieurs décennies, etc. - un modèle de pattern doit représenter les données de manière à construire une image informative et fidèle de cette réalité sur laquelle nous avons des données. On attend de cette image qu'elle puisse être mathématiquement représentée sous une forme quelconque – ce qui, entre autres, permet extrapolations et prédictions. Si les données sont souvent des points dans un espace, le modèle de pattern sera le graphe d'une fonction

mathématique qui ajuste au mieux ces points. De manière plus générale, toute corrélation entre séries de données est un pattern, mais les corrélations (positives ou négatives) fortes sont les seules susceptibles de motiver l'élaboration d'un modèle mécaniste (qui répondrait à la question de la raison de cette corrélation).

Le processus, lui, est censé produire le pattern. Le modèle de processus, ou modèle mécaniste, représentera le comportement d'un système, ou d'entités, susceptibles de produire le pattern considéré. Ainsi, si on met en évidence un pattern d'augmentation exponentielle du taux de CO<sub>2</sub> dans l'atmosphère, le modèle mécaniste représentera le fonctionnement un système d'activités sociales, économiques et écologiques plausibles qui entraînent une telle augmentation. Il est à noter que les questions d'identification de facteurs causaux se posent précisément dans ce contexte : après avoir trouvé un pattern de corrélation entre consommation de tabac et cancer du poumon, on élaborera un modèle mécaniste des processus biochimiques d'interaction entre tabac et organismes qui nous dira s'ils peuvent augmenter la probabilité de développer un cancer du poumon, et ainsi voir la corrélation comme une relation causale.

En particulier, la question de l'efficace causale d'un facteur donné met souvent en jeu la relation entre modèle de processus et modèle de pattern dans ce qu'on appelle « test d'hypothèse ». Sans entrer dans les complications innombrables de ce protocole (Fisher, 1935, Graves et Gotelli, 1996, Burnham et Anderson, 2002) il arrive que l'on compare le pattern existant, une fois modélisé, avec le pattern qui serait obtenu par un processus aléatoire, autrement dit, un processus dans lequel le facteur causal que l'on veut tester ne joue aucun rôle – ce que l'on nomme « hypothèse nulle ». Si la différence est statistiquement négligeable, alors cela signifie que le processus hypothétiquement imaginé ne joue en réalité aucun rôle majeur dans le phénomène caractérisé par le pattern modélisé.

Ces distinctions sont souvent contextuelles ; ainsi, la loi des gaz parfaits ( $PV = nRT$ ) peut être un modèle de pattern lorsqu'on fait de la cinétique des gaz (et le modèle mécaniste modélisera alors les interactions entre particules qui donnent lieu à cette loi) aussi bien qu'un modèle de processus lorsque par exemple la théorie porte sur le fonctionnement d'un moteur, et que le comportement d'un gaz y entre comme un « ingrédient ». Néanmoins, cette distinction, présente dans la langue vernaculaire des scientifiques, est bien utile.

### **Modèles de pattern : mathématiques et réalité.**

La mise au jour de patterns – spatiaux, temporels, génomiques, etc. - est un aspect majeur des sciences biologiques. Ainsi du côté de la biologie évolutive, la paléontologie se concentre sur l'élaboration de modèles de patterns. La distribution et la forme des clades à l'échelle macro-évolutive est un pattern important qui a donné lieu à de nombreuses controverses (figure 1). La question de la variation de la complexité (quoi qu'on entende par ce terme) ou bien de la taille des organismes à travers le temps se pose à partir du moment où on a identifié comme pattern à travers le registre fossile une variation dirigée significative de ces quantités (Mc Shea 1996, Turner 2011) (figure 2). En

écologie, la relation entre superficie et nombre d'espèces, ou bien la distribution de la biodiversité en termes de nombres d'espèces par effectif donné (*species abundance distribution*) sont des objets essentiels, car c'est seulement une fois qu'on a modélisé ces patterns qu'on peut se faire une idée de la biodiversité existante, et qu'on peut ensuite s'interroger sur les mécanismes qui la créent et la maintiennent (Figure 3).

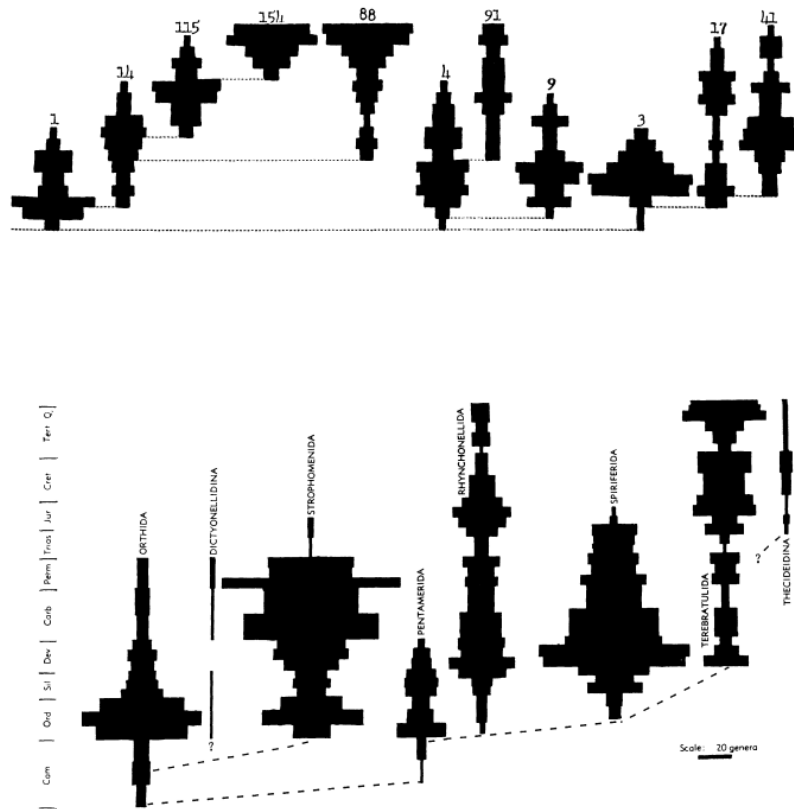


Figure 1. Comparaison de l'évolution des clades reconnus dans le registre fossile (bas) et d'un pattern généré par un modèle (haut) – d'après Gould et al. 1977

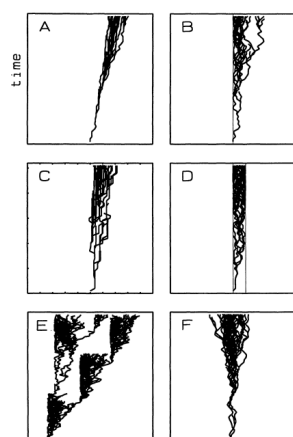
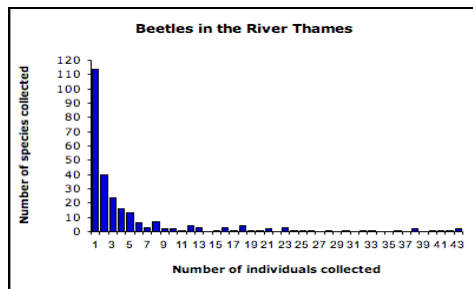


Figure 2. Plusieurs types de pattern d'une tendance à la complexité dans la phylogénèse, générés par des modèles faisant des hypothèses différentes sur les processus évolutifs en jeu (Mc Shea 1996)

a.



b.

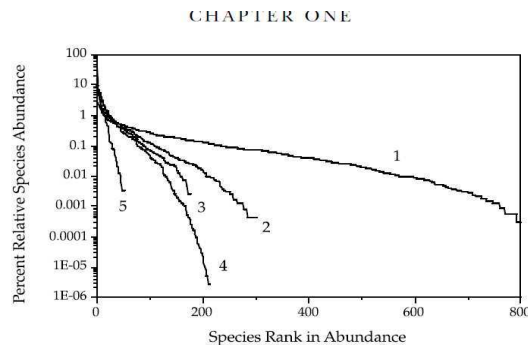


FIG. 1.1. Patterns of relative species abundance in a diverse array of ecological communities. Species in each community are ranked in percentage relative abundance from commonest (*left*) to rarest (*right*). The percentage relative abundance is log transformed on the y-axis. 1: Tropical wet forest in Amazonia. 2: Tropical dry deciduous forest in Costa Rica. 3: Marine planktonic copepod community from the North Pacific gyre. 4: Terrestrial breeding birds of Britain. 5: Tropical bat community from Panama.

Figure 3. Distributions d'abondance d'espèce. a. Sous forme d'effectifs ; b. sous forme de rang.

Les modèles de pattern présentent des difficultés spécifiques, que nous signalons brièvement :

- *Overfitting* versus prédiction.

Un bon modèle de pattern doit pouvoir placer au mieux les données sur les courbes mathématiquement définies qu'il représente (*curve-fitting*, en anglais). Mais au fond, la courbe qui colle le mieux aux données permet assez peu de prédire les prochaines données ; à l'inverse, chercher une courbe simple permet d'avoir un bon pouvoir prédictif, mais au prix d'une perte de proximité avec les données. Ainsi, dans la figure 4, tandis que  $h(x)$  - la courbe la plus simple (linéaire) - ne colle pas du tout aux données,  $g(x)$  - la parabole - est moins simple mais beaucoup plus simple que la ligne brisée représentée sur 4 b, qui est une suite de fonctions linéaires, et qui ne permet pas du tout de prédire le prochain point : à trop coller aux points, le pattern ne peut plus rien prédire. Un modèle de pattern comme sur cet exemple, réalise donc toujours un compromis entre qualité de l'ajustement et simplicité.

- Simplicité

Mais la simplicité elle-même est une notion délicate. Comme le rappelle Sober (2002) supposons deux fonctions,  $y = ax^2 + bx + c$ , et  $y = bx + c$ . Intuitivement, on dira que la seconde est plus simple car elle a moins de paramètres, mais on peut tout aussi bien soutenir que la première est

plus simple, car la seconde est un cas particulier de la première, dans lequel on a assigné à  $a$  la valeur 0 (information supplémentaire, qui la rend moins simple). Il est donc assez difficile de définir la simplicité d'une courbe (donc d'un modèle de pattern) dans l'absolu. A la difficulté définitionnelle s'ajoutent des complications épistémologiques et métaphysiques concernant les raisons pour lesquelles la simplicité devrait être (ou non) préférée.

- Calibration et confirmation.

Duhem dans *La théorie physique, son objet, sa structure* a souligné le fait qu'un même jeu de données peut être représenté par une infinité de fonctions mathématiques différentes. Ainsi le compromis que nous évoquions – entre qualité de l'ajustement et simplicité – est la réponse pragmatique des scientifiques à cette question de principe, et ceci vaut en biologie comme en physique (dans les deux cas on reconstitue des patterns). Cela implique aussi que la vérification d'un modèle de pattern enveloppe en fait deux opérations distinctes, la calibration et la confirmation. Si j'ai des données sur un phénomène, je peux certes le représenter par plusieurs courbes différentes : mettons, une fonction hyperbolique -  $y = g(x) = 1/cx^d$  - et une fonction exponentielle -  $y = f(x) = \exp(-ax) + b$ . On appellera « confirmation » l'opération qui consiste à comparer à des nouvelles données une fonction déjà définie et paramétrée (par exemple,  $a$  et  $b$  sont fixés dans l'exponentielle) et prise pour modèle phénoménologique de données existantes, afin de confronter celle-ci à ses prédictions et de la tester de manière hypothetico-déductive. Mais on peut aussi poser d'abord que la courbe est en général d'une certaine nature (par exemple hyperbolique) et ensuite, la comparer aux données afin de préciser les paramètres : il s'agit alors d'une « calibration ». Le processus de vérification pose des problèmes spécifiques (par exemple, se sert-on vraiment deux fois de suite des mêmes données, une fois dans la confirmation puis une fois dans la calibration, ce qui poserait à la vérification des modèles un problème de circularité épistémique ? - cf. Werndl 2013). Mais surtout, il est clair que si l'on a des raisons *indépendantes* de penser qu'il existe dans la réalité un processus susceptible de produire un pattern de telle ou telle allure – par exemple, une fonction exponentielle – alors on optera pour cette fonction, et ensuite on la calibrera. La validation de modèle de pattern n'est donc jamais indépendante d'informations que l'on peut avoir sur les processus susceptibles de produire le pattern. Il y a là comme une circularité méthodologique: on cherche un modèle mécaniste à partir du pattern ; mais toute information indépendante sur de possibles mécanismes est utilisée pour choisir le type de fonction/ courbe.



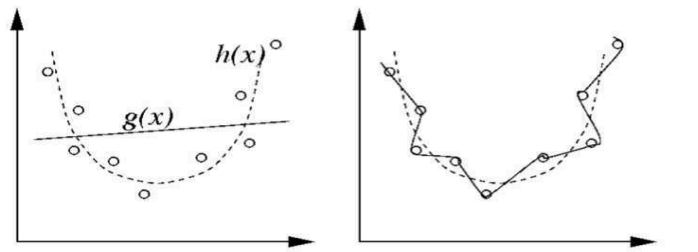


Figure 4. *Overfitting et simplicité :  $g(x)$  est trop simple, la courbe en trait plein de la figure de droite colle adéquatement à toutes les données mais  $h(x)$  réalise un meilleur compromis entre pouvoir prédictif et adéquation.*

Une fois constatée cette non-indépendance des modèles de processus et des modèles de patterns, nous nous tournons maintenant vers les modèles de processus en biologie.

### **Modélisation mathématique de processus.**

*Modèles de processus en biologie évolutive.*

*L'évolution par sélection naturelle et ses conditions formelles.*

Darwin (1859) a affirmé deux choses : l'arbre de la vie – toutes les espèces existantes descendent d'espèces anciennes, et la proximité taxonomique est donc toujours une proximité historique ; et cet arbre de vie - donc la diversité des vivants, comme leurs adaptations - résultent essentiellement de la sélection naturelle. Ces deux thèses constituent le cœur de la biologie évolutive, dont à l'instar de l'adage célèbre de Dobzhansky – « rien en biologie n'a de sens excepté à la lumière de l'évolution » - il est raisonnable de penser qu'elle fournit un cadre/horizon incontesté à l'ensemble des théories biologiques. Néanmoins, on a souvent remarqué que Darwin n'adopte aucun outil mathématique.

Ernst Mayr (1976) écrit que Darwin a substitué à un mode de penser « typologique », pour lequel chaque individu est simplement un représentant du type de son espèce, une pensée « populationnelle », dans laquelle les différences interindividuelles sont fondamentales, donc pour laquelle les populations d'individus (et non les individus eux-mêmes) sont épistémologiquement cruciaux. La sélection naturelle est en effet un phénomène fondamentalement populationnel, puisque son cœur est la reproduction différentielle : certains individus se reproduisent plus que d'autres du fait de certaines de leurs propriétés héréditaires (Huneman 2009). Ce mode de pensée populationnel essentiel à la biologie évolutive nécessite les mathématiques statistiques et la théorie des probabilités s'il doit donner lieu à des modèles des processus en jeu.

Darwin n'avait pas de théorie correcte de l'hérédité, mais il a montré que des variations héréditaires sont nécessaires pour que la sélection naturelle ait lieu. Mendel a fourni une conception plus adéquate, dans laquelle les déterminants des traits héréditaires sont des entités discrètes appelées « gènes », et Fisher comme Wright ont montré que cette conception particulière (ou discrète) de l'hérédité rendait possible l'évolution par sélection naturelle beaucoup mieux qu'une conception de l'hérédité comme « mélange » continu de

qualités du père de la mère (Gayon 1992). Comme l'a montré Lewontin (1970) cette évolution a lieu dès qu'existe une population d'individus pourvus de traits *variables, héritables*, et qui influencent les *chances de reproduction* de leur porteur (voir Heams 2009 a, b sur hérédité et variation, Merlin 2013 sur la variation).

Depuis Fisher, une variable nommée « fitness » nomme ces chances de reproduction différentielle (même si les controverses font rage sur l'interprétation exacte de la fitness, voir Rosenberg et Bouchard 2010). Les généticiens des populations ont montré mathématiquement que dans de grandes populations un avantage en fitness finit par garantir à son bénéficiaire une fixation dans la population.

L'héritabilité est, elle, avant tout un phénomène statistique : la valeur d'un trait chez les descendants d'individus tend à se comporter comme la valeur de ce trait chez les parents par rapport à la moyenne (les grands ont tendance à faire des enfants plus grands que la moyenne). Ainsi conçu le phénomène de l'héritabilité se laisse modéliser *statistiquement*. L'héritabilité mesure avant tout la partie de la *variance* d'un trait qui dépend de la transmission intergénérationnelle (donc généralement génétique, mais aujourd'hui ce point est aussi controversé, voir Danchin et al. 2011) et non de l'environnement.

*Génétique des populations et génétique des quantitatives, cœur de la Théorie moderne de l'évolution – et leurs modèles mathématiques.*

La théorie dite synthétique de l'évolution (TSE), initiée par la génétique des populations (Fisher, Wright, Huxley, Haldane..) dans les années 30, a produit une compréhension mathématique du processus de sélection naturelle, et donné à la biologie de l'évolution sa forme mathématique actuelle (Gayon 1992). Cette compréhension se fonde avant tout sur la théorie mathématique des probabilités et des statistiques, Fisher et Wright ayant d'ailleurs largement contribué aux avancées de cette dernière.

La génétique des populations étudie la variation des fréquences géniques dans les populations ; développée dans le même contexte, la génétique quantitative (Falconer 1961) étudie la variation de la valeur de traits à grandeur continue (tels que la taille) déterminés par de nombreux gènes (Wright 1932, Fisher 1930). Ces études permettent de saisir l'effet de la sélection naturelle : en effet, si un trait influence positivement la reproduction de son porteur à la génération suivante, il a de fortes chances d'augmenter en fréquence, et sa valeur se modifiera sous l'effet de cette transformation intergénérationnelle répétée. Un modèle mathématique de l'effet cumulé de cette variation permet de saisir le processus de sélection naturelle imaginé par Darwin, tel qu'il s'exerce sur un grand nombre de générations. A partir de là les auteurs de la TSE pensent que l'on peut comprendre le processus d'évolution par sélection naturelle en tant qu'il donne lieu à des changements d'organismes, d'espèces, de familles.

Plusieurs types de modèles permettent alors de décrire cette évolution par sélection naturelle. Ils utilisent différents outils mathématiques pour capturer le rapport entre fitness des traits ou des génotypes, constitution génétique de la population et changements consécutifs de valeur des traits. Ainsi, en génétique quantitative, une équation très simple (dite *breeder's equation*)

décrit ce rapport : le changement intergénérationnel dû à la sélection est le produit de l'héritabilité et de l'intensité de la sélection (que nomme une variable appelée « coefficient de sélection »). Des modèles plus sophistiqués décomposent l'héritabilité en termes de variances génétique additive et de covariances des différents traits en jeu ; ainsi, l'équation devient une équation vectorielle  $B = G B$  (le vecteur  $B$  étant constitué du changement valeur de  $x$  traits entre les générations  $n$  et  $n-1$ , et  $R$  étant constitué des coefficients de sélection pour chaque trait) dans laquelle la matrice (nommée matrice- $G$ ) représente la structure génétique de la population par ces variances et covariances entre traits (Lande et Arnold 1983).

En génétique des populations, des modèles stochastiques décrivent l'évolution de la population en termes probabilistes ; les modèles décrivent souvent l'évolution sur un, ou parfois deux, loci, d'une population avec un effectif  $N$  et différentes particularités de structure (accouplement aléatoire ou pas, castes etc.), mettant en jeu de nombreux allèles sur ce ou ces loci, et pourvues de taux de mutation et de migration donnés (Gillespie 2004). On peut partir d'un équilibre dit « de Hardy-Weinberg » donné par les fréquences géniques dans une population de diploïdes en cas d'absence de sélection, migration et mutation (avec assortiment au hasard et population infinie). Cet équilibre donné par une formule simple issue des lois de Mendel est stable au bout d'une génération comme démontré par Hardy et Weinberg ; on introduit alors les valeurs de fitness des différents allèles ou génotypes en tant qu'elles affectent la probabilité de transition des fréquences géniques d'une génération à l'autre, et on peut estimer les modifications probables à court et long terme des fréquences génotypiques ou allélique de la population.

Différents modèles réalisent différentes assumptions sur les populations: avec ou sans générations enchevêtrées, type d'accouplement, etc. Le modèle de Wright-Fisher, le premier et un des plus classiques, permet en particulier - à partir de la simplification raisonnable consistant à modéliser la probabilité de ré-occurrence d'un allèle à la génération suivante par une loi de distribution binomiale - de modéliser les effets de l'erreur d'échantillonnage dans les petites populations, mis en évidence par Wright et appelé « dérive génétique aléatoire ». De même que le modèle ultérieur de Moran (à générations enchevêtrées), ce dernier est un modèle Markovien : chaque état ne dépend que de l'état immédiatement précédent, et l'évolution, toujours stochastique, est donnée par les fonctions de transition. On y calcule les probabilités de fixation à un temps  $T$  donné des différents allèles sous différentes conditions initiales de population, avec des probabilités de transitions, supposées égales pour chaque allèle quand il n'y a pas de sélection.

Comme y a insisté Fisher (1930), ces modèles présentent des analogies avec ceux de la mécanique statistique : dans les deux cas, on considère une population d'un grand nombre d'éléments (des molécules, des gènes) en interaction, et on construit des variables qui décrivent des ensembles d'états individuels : entropie, fitness d'un trait (calculé comme probabilité de descendants d'une classe individus partageant un trait). L'identité formelle entre mécanique statistique et génétique de populations a été justement explicitée dans le cas de la dérive génétique (identifiable à une marche aléatoire) (Blythe and Mc Cann 2010) ; certains auteurs (e.g. Barton and Coe 2009, Sella and Hirsh 2005) ont tenté de démontrer une identité générale entre mécanique statistique

et génétique des populations ou quantitative en construisant la fitness comme un certain type d'entropie, et en démontrant des équations très générales pour tous ces systèmes, mais cette formalisation, extrêmement technique, n'a que peu d'usage en biologie.

Quoi qu'il en soit, génétique des populations et génétique quantitative peuvent donner lieu à des équations très générales qui décrivent l'évolution par sélection naturelle : outre le *breeders equation* et ses sophistications récentes, Price (1970) a montré comment le changement évolutif intergénérationnel pouvait être écrit comme une addition de deux termes, une covariance et une espérance :

$$Dz = \text{Cov}(w_i, z_i) / w + E(w_i Dz_i) / w$$

( $z$  est la valeur moyenne du trait,  $w$  la fitness moyenne,  $w_i$  la fitness de l'individu  $i$  et  $z_i$  sa valeur pour le trait  $z$ ,  $Dz_i$  est le changement de cette valeur entre un individu et son descendant).

Le terme de covariance capture l'effet de la sélection, l'autre capture le biais dans la transmission du trait. Cette équation très formelle, qui ne dit rien sur les causes du changement évolutif puisqu'elle n'est qu'une réécriture de la description statistique de la différence entre deux générations, permet aussi de décomposer la sélection naturelle en sélection au niveau des individus / sélection au niveau des groupes (Okasha 2006) – elle est souvent utilisée dans ce contexte.

D'autres formulations très générales de l'évolution ont été fournies, dans des contextes mathématiques différents : l'équation des réplicateurs (en dynamique des réplicateurs, une forme de théorie évolutionniste des jeux (Nowak 2006)) ou bien l'équation dite « équation canonique » dans la théorie dite « dynamique adaptative » - qui conçoit la fitness d'un individu (en génétique des populations) ou d'une espèce (en écologie) comme probabilité d'invasion d'un mutant (Metz 2008).

#### *Écologie comportementale : les mathématiques de l'adaptation.*

Si en modélisant le processus d'évolution par sélection naturelle la génétique des populations constitue pour la TSE le cœur de notre connaissance de l'évolution, elle n'épuise pas l'évolution biologique, en particulier elle ne nous dit pas quelles sont les fonctions des différents traits sélectionnés, c'est-à-dire (Walsh 1996) les raisons de leur fitness plus élevée. L'écologie comportementale (Krebs et Davis 1997), élaborée à la suite de l'éthologie dans les années 1970-80, vise à expliquer pourquoi certains traits sont sélectionnés plutôt que d'autres, autrement dit, leur rôle adaptatif dans des environnements donnés. Elle commence par postuler les traits présents comme étant sélectionnés, donc comme maximisant la fitness dans leur environnement, et les conçoit ainsi comme des réponses aux demandes environnementales (prédateurs, température, rareté des ressources, etc.). Mathématiquement, il s'agit donc de modéliser le rapport entre valeur d'un trait phénotypique et fitness – ou représentant de la fitness, tel que apport énergétique, vitesse de métabolisme, etc., bref tout trait dont il est raisonnable de penser que dans un environnement donné il est directement corrélé à la fitness. Dans ce cas, le biologiste use d'équations différentielles pour représenter ce rapport, et le calcul d'extrema

indique les valeurs de trait qui sont adaptatives (si la réalité ne correspond pas à cet optimum prédit, on cherche d'autres demandes environnementales, ou des contraintes qui expliquent l'impossibilité de l'atteindre, cf. Maynard Smith et al. 1985).

Cependant certains traits – par exemple l'agressivité, l'altruisme, le mimétisme animal – dépendent de leur fréquence chez les individus avec lesquels l'individu focal interagit : si l'agressivité semble un trait profitable dans un monde de peureux, dans un monde d'agressifs, il vaut mieux être peureux et éviter le combat qu'être agressif. Dans ce cas, il n'y a pas d'optimum (puisqu'une fois atteint par tous, l'optimum n'en est plus un) mais on vise à comprendre les stratégies les plus stables, celles dont la fréquence ne diminue pas dès qu'une stratégie alternative apparaît ; ces stratégies dites « évolutionnairement stables » ou SES (Maynard Smith 1982) sont l'objet de modèles de théorie des jeux. Ici mathématiquement la biologie évolutive ne voisine donc plus avec la physique statistique, mais avec la microéconomie (ces parentés sont développées dans Grafen 2007, Huneman 2013, Martens 2014).

#### *Ecologie des communautés et modèles de systèmes complexes.*

Si l'écologie comportementale étudie les individus, ce qu'on nomme habituellement « écologie », soit l'écologie des populations, des communautés ou des écosystèmes, étudie des ensembles interactifs d'individus, parfois de différentes espèces (écologie des communautés). En écologie des populations l'un des plus anciens et célèbres modèles est le cycle de Lotka-Volterra, qui décrit les variations d'effectifs d'une population de proies et de prédateurs (encadré 1). Ce sont des équations différentielles qui représentent la variation de l'effectif de chacune des espèces en fonction des taux de morts et naissances de l'autre, formalisant le fait que les prédateurs dépendent de l'effectif de proies pour subsister, et en retour, l'effectif de proies à chaque génération paye une dette proportionnelle à l'effectif des prédateurs. Sous une autre forme elle peut modéliser la compétition.

##### Encadré 1. Les équations de Lotka-Volterra.

$$dx(t)/dt = \alpha x - \beta xy$$

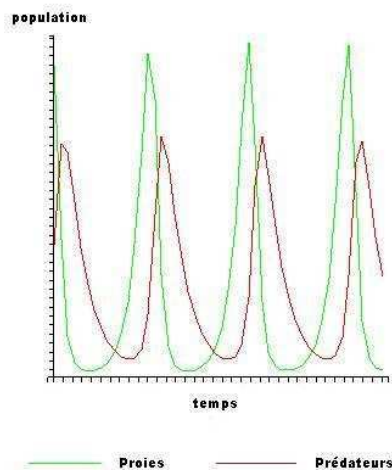
$$dy(t)/dt = -\gamma y + \delta xy$$

où  $t$  est le temps ;  $x(t)$  est l'effectif des proies en fonction du temps ;  $y(t)$  est l'effectif des prédateurs en fonction du temps ;  $\alpha$  est le taux de reproduction des proies (constant, indépendant du nombre de prédateurs) ;  $\beta$ , taux de mortalité des proies dû aux prédateurs rencontrés ;  $\gamma$ , taux de mortalité des prédateurs (constant, indépendant du nombre de proies) ;  $\delta$ , taux de reproduction des prédateurs en fonction des proies rencontrées et mangées.

Ces équations décrivent des oscillations, et ont permis de montrer que sous certaines conditions la lutte contre une espèce entraîne la disparition de l'autre (figure 5). Le cycle de Lotka-Volterra atteint un niveau extrême de généralité

puisque l'on ne précise rien hormis les taux de naissance et de mort de chaque espèce. Elles peuvent être généralisées sous une forme vectorielle pour  $n$  espèces, et décrire des relations de compétition comme de prédation entre elles (et correspondent à l'équation des répliqueurs sous certaines conditions). Le comportement de ce système peut souvent être intrinsèquement chaotique (au sens de dépendance extrême aux variations de conditions initiales).

Figure 5. Cycles de Lotka Volterra.



Il est clair aussi que souvent, les ensembles d'équations différentielles qui décrivent l'évolution d'un système en génétique des populations ou en écologie - parfois bien plus complexes que celles de Lotka-Volterra - sont mathématiquement intraitables, soit en général, soit même dans la plupart des cas particuliers ; on a alors recours à des modèles qui simulent, via des algorithmes reproduisant de manière discrète le comportement des équations, l'évolution des systèmes considérés. Le choix du type de modélisation, et les différences, identités et intérêts spécifiques des équations vs. algorithmes, sont des questions méthodologiques discutées et cruciales (e.g. Gaucherel et al 2011). Les simulations - que ce soit modèles à base d'agents, algorithmes génétiques (Schoenauer 2009) ou automates cellulaires - sont alors très utilisés en biologie évolutive ou en écologie (par exemple le modèle Echo de Holland) (Huneman 2014). L'écologie des écosystèmes et les sciences de la Terre utilisent de tels modèles pour s'interroger sur le genre de conséquences que peut avoir l'influence de la vie sur la dynamique de l'environnement et répondre à des questions contrefactuelles (comme : à quoi ressemblerait l'environnement planétaire si la vie ne l'influait pas ?) - le modèle dit Daisyworld, modélisant la vie par deux types de marguerites, est le plus connu (Watson et Lovelock 1983, Dutreuil 2014).

Les modèles de processus en écologie posent de profondes et nombreuses questions épistémologiques, qui suscitent entre écologues de vifs débats. Richard Levins - un écologue dont le travail a contribué à importer le cadre de la TSE dans l'écologie (définie autrefois par Haeckel comme « science de la lutte pour la vie ») est aussi l'auteur d'une réflexion épistémologique qui a durablement marqué l'écologie, et plus généralement la philosophie des sciences. Dans « The

strategy of model building in ecology » (1966), il soutient en effet que des modèles en écologie – et sans doute au-delà, en tout cas en biologie – ne peuvent à la fois satisfaire des valeurs de généralité, de réalisme et de précision (ou prédictibilité). (Sa réflexion concerne avant tout les modèles de processus, mais on peut sans doute l'adapter pour les modèles phénoménologiques). Ainsi le système Lotka-Volterra est très général, mais on ne peut pas le rendre plus réaliste (par exemple en intégrant des paramètres décrivant des espèces particulières) sans perdre en généralité, ou sinon en précision. Les exemples que donne Levins concernent en particulier des modélisations d'un même problème (fitness en environnement variable) en termes purement phénotypiques (et donc très généraux par rapport aux caractères d'espèces particulières) ou bien en termes de fréquences génotypiques (ce qui est plus réaliste puisque cela ne concerne que des espèces partageant une constitution génétique particulière). Il en déduit d'une part que chaque modèle mathématique doit opérer des compromis entre ces trois valeurs épistémiques, de l'autre que seules sont susceptibles de saisir la réalité des propositions obtenues dans plusieurs modèles ainsi distincts. Cette réflexion indique une voie pour comprendre l'apport de la modélisation mathématique à l'explication de la réalité biologique.

### *Modèles de processus en biologie fonctionnelle.*

La biologie « fonctionnelle » définie par Mayr est sûrement plus hétérogène que la biologie évolutive puisqu'elle manque d'une référence commune équivalente à la référence commune que font toutes les disciplines évolutives au processus de l'évolution. Biologie cellulaire, physiologie, biologie moléculaire, immunologie, ou biologie du développement – chacune étant une instance de biologie fonctionnelle – ne s'intègrent donc pas dans un cadre théorique général ; les modèles mathématiques que chacune emploie sont donc très variés, mais certains leur sont communs. Nous en présentons quelques exemples, pour souligner la valeur de généralité de certains cadres mathématiques.

La biologie moléculaire, sans doute l'un des domaines les plus développés de toutes les sciences de la vie (en nombre de publications, de chercheurs ou de financement) a depuis ses origines, avec la découverte de l'ADN et du code génétique, promu la notion d'information génétique. Or il existe depuis les années 30 une théorie mathématique de l'information, élaborée par Shannon dans le cadre de problèmes de transmission de messages par des systèmes de communication ; l'information de Shannon donne lieu à une mesure, définie en termes de probabilités. Si les biologistes moléculaires parlent facilement d'information lorsque qu'ils exposent leurs théories, dans l'élaboration même des modèles de biologie moléculaire il semble que les concepts mathématiques d'information jouent peu ou pas de rôle (au contraire de concepts verbaux comme « signal », « mémoire », « code » etc.). En outre, les différents usages de l'information en biologie ne se laissent pas toutes embrasser par la théorie mathématique de Shannon : des controverses ont cours concernant à la fois la légitimité biologique du langage informationnel comme tel (e.g Griffiths 2001), l'univocité de la notion d'information en biologie, et, le cas échéant, le cadre mathématique qui serait propre à en rendre compte.

Notre premier exemple de modèle de processus en biologie fonctionnelle vient de la biologie du développement, dont le problème majeur consiste à comprendre la différenciation : différenciation de la cellule originaire en cellules distinctes même si génétiquement homogènes, formation de formes organiques différenciées au cours de l'embryogénèse (e.g. création de feuilles et disposition spécifique de celles-ci sur l'axe de la tige, création d'organes complexes). Alan Turing, le mathématicien connu pour sa contribution majeure à l'informatique, a dans l'article de 1952, *The chemical basis of morphogenesis*, élaboré un modèle mathématique dit « modèle de Turing », que nous présentons brièvement ici pour son intérêt.

Alors que les travaux mathématiques antérieurs de d'Arcy Thompson sur la forme biologique étaient essentiellement *descriptifs* et consistaient à montrer des analogies structurelles entre des phénomènes mécaniques et des formes biologiques ou à montrer que des structures mathématiques sous-tendaient l'organisation des êtres vivants, de même qu'était *descriptive* une partie importante de la biologie du développement du début du 20<sup>e</sup> siècle, ce que Turing propose est un modèle générique exposant des *mécanismes* permettant de rendre compte de la création des formes biologiques.

Plus spécifiquement, le problème est formulé en termes de perte de symétrie: supposant qu'un embryon possède initialement une symétrie sphérique, la question importante qui se pose pour Turing à propos de la morphogénèse concerne la nature du mécanisme qui permet à cet embryon de *perdre* des degrés de symétrie (pour former, par exemple, un animal comme un cheval qui ne possède plus la symétrie initiale de l'embryon).

L'explication de Turing repose sur le couplage de la *diffusion* et de la *réaction* de molécules chimiques que Turing propose de nommer « morphogènes ». Les processus de diffusion sont décrits suivant les équations standard de diffusion, connues depuis le 19<sup>e</sup> siècle, décrivant la migration des molécules des régions les plus concentrées vers les moins concentrées, à une vitesse proportionnelle au gradient de concentration et à un facteur spécifique à la substance (diffusibilité,  $D$ ). A ces équations, Turing ajoute un terme décrivant la réaction des molécules suivant la loi d'action de masse (la vitesse de réaction est proportionnelle à la concentration de la substance) : la diffusion seule, tendant à homogénéiser le milieu, ne pourrait en effet suffire à établir des hétérogénéités.

Turing montre ainsi que le simple couplage de ces deux processus (réaction et diffusion) peut suffire à rendre compte de la perte de symétrie. Etablissant une analogie avec les phénomènes de résonance, bien connus en mécanique ou en électricité, il montre que des variations aléatoires initialement très faibles au sein de l'embryon peuvent être amplifiées et donner lieu à des comportements instables ou stables mais loin de l'équilibre (le système n'est donc initialement pas *strictement* homogène, on doit s'autoriser des perturbations initiales aléatoires faibles). La production de deux morphogènes, X et Y, peut ainsi, dans certaines conditions, mener à la création de patterns spatiaux qui brisent la symétrie initiale. C'est notamment le cas dans une situation où la diffusibilité de Y est grande devant celle de X, X catalyse sa propre production, X réagit pour être transformé en Y et Y inhibe la production de X.



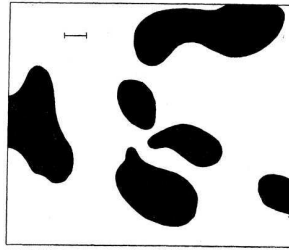


Figure 6 - Exemple de pattern tacheté proposé par Turing, correspondant à la première possibilité évoquée dans le texte pour un modèle de  $N$  cellules en anneau.

Le modèle que Turing considère le plus longuement est celui d'un anneau fermé constitué de  $N$  cellules et dans lequel 2 (ou 3) morphogènes sont produits. L'examen du modèle montre que six possibilités peuvent émerger. Pour une de ces possibilités, Turing indique qu'elle peut rendre compte des motifs tachetés que l'on observe chez les animaux (figure 6). Dans une autre de ces possibilités, on observe un pattern (invariant dans le temps) de variation de la concentration des morphogènes par vague au sein de l'anneau (figure 7). Pour Turing, ce pattern peut rendre compte de la disposition des feuilles autour de la tige d'une plante aussi bien que de celle des tentacules autour de l'axe d'un hydre ; l'anneau de cellules considéré dans le modèle correspond à une section de la tige (ou de l'axe) et les régions de l'anneau plus riche (ou plus pauvre, peu importe) en tel morphogène peuvent correspondre au lieu d'initiation des feuilles ou tentacules.



Figure 7 – Hydre d'eau douce (à gauche) et aspérule odorante (à droite). Pour les deux espèces un motif donné (une tentacule, resp. une feuille) est initié selon une période régulière autour d'un anneau de cellules (la terminaison de l'hydre, une section de la tige de l'aspérule).

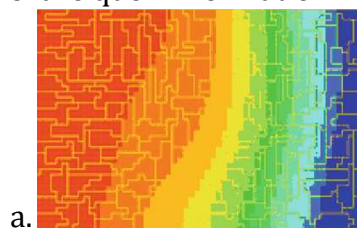
Il est remarquable que le même modèle (considérant un anneau de cellules) puisse expliquer des phénomènes apparemment aussi différents que la phyllotaxie de plantes comme l'Aspérule odorante et la disposition des tentacules chez l'Hydre, puis qu'une variante du modèle (considérant une sphère) puisse rendre compte de la gastrulation (phase correspondant à la perte de symétrie sphérique lors du développement des métazoaires). La diversité des phénomènes qu'il propose d'expliquer, comme la manière dont il transcende une division pourtant aussi marquée que l'est celle entre biologie du développement animal (Métazoaire, Hydre) et biologie du développement végétal (Aspérule odorante), classent le modèle de Turing dans le pôle le plus générique de la classification de Levins.

L'article de Turing a eu une influence tout à fait décisive en biologie du développement ; il a par ailleurs inspiré les travaux ultérieurs de René Thom, lui aussi mathématicien s'étant penché sur la morphogenèse. Les premières

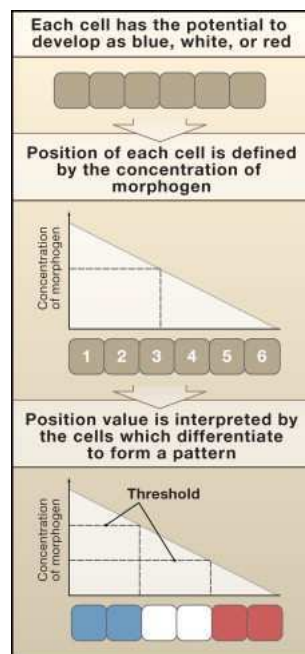
confirmations expérimentales du modèle considérant un anneau de cellules continu, c'est-à-dire non compartimenté, sont intervenues à la fin des années 1980 ; celle du modèle considérant un anneau de cellules compartimenté, en 2014 (voir Tompkins *et al.* 2014 et les travaux qui y sont cités).

Les modèles de réaction-diffusion sont courants en physique ; on soulignera qu'ils permettent aussi, par les mêmes équations mathématiques, de modéliser des phénomènes écologiques (Colizza *et al.* 2007) ou épidémiologiques ; en sociologie aussi, des phénomènes tels que la propagation de la délinquance ou des modes obéissent à des équations de réaction-diffusion (Short *et al.* 2010).

Si la formation de structures a été éclairée par le modèle de Turing, la différenciation cellulaire, autre problème majeur de la biologie du développement, a été abordée à partir des travaux de Lewis Wolpert dans les années 60 dans le cadre du modèle dit du Drapeau Français (*french flag model*) (Figure 8); dans ce modèle, chaque cellule d'un organisme multicellulaire peut adopter de nombreux phénotypes (cellule de muscle, de nerf etc.) et va « choisir » son type en fonction d'une information, dite « information de position » donnée par le gradient d'une substance morphogène émise par les cellules avoisinantes. (A noter que cette information de position n'est pas du même ordre que l'information « génétique » de la biologie moléculaire.)



a.



b.

Figure 8. Modèle du drapeau français en théorie du développement. a) pattern de drapeau français propre au gradient de morphogène ; b) mécanisme sous-jacent (d'après Kerszberg *et Wolpert* 2007)

Notre dernier exemple concerne la biologie moléculaire. Ses avancées, et les outils techniques et computationnels permettant l'analyse de l'expression de génomes entiers, ont suscité l'introduction de nouveaux outils mathématiques pour comprendre l'expression du génome. Le produit d'un gène est en effet conditionné par les interactions innombrables d'un ensemble de gènes, et d'éléments présents dans les cellules (ARN, micro-ARN, protéines, etc.). Ce système d'interactions peut être représenté par des réseaux, objet qu'étudie la théorie des graphes (de la Fuente 2009). Un réseau est un ensemble de nœuds, reliés entre eux par des arêtes. Le degré de connexion de chaque nœud avec les autres, le nombre de connections nécessaires pour aller d'un nœud à un autre, etc., constituent des propriétés mathématiques du réseau, qui présente alors des propriétés qu'analyse la théorie des graphes (Strogatz 2001). L'approche par graphes est devenue fondamentale à plusieurs niveaux de la réalité biologique : le génome donne lieu à des « Gene regulatory networks » (GRN) (Davidson 1986) qui permettent de comprendre l'expression des gènes et plus généralement le fonctionnement du génome (fig. 9) ; la biologie cellulaire étudie des réseaux métaboliques d'interaction entre les produits cellulaires. Les réseaux permettent d'identifier comme sous-parties structurées des « motifs » récurrents (Alon 2007), qui peuvent avoir un rôle fonctionnel et incitent à étudier leur conservation et leurs transformations durant l'évolution.

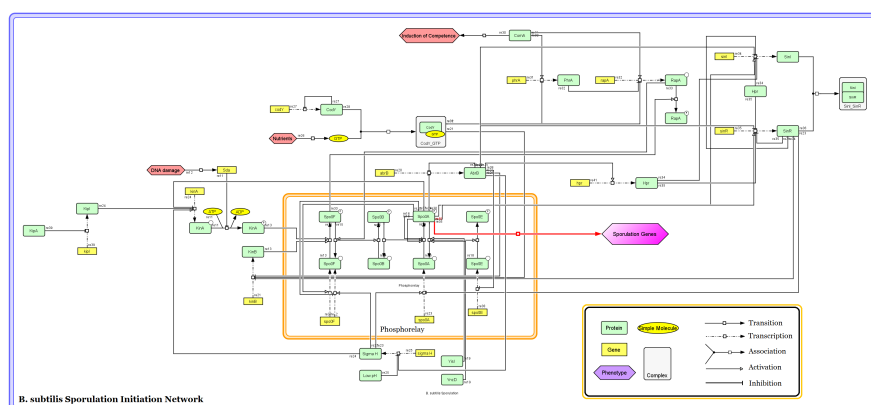


Figure 9. GRN pour l'initiation de la sporulation chez *Bacillus subtilis*.

Les GRN ont la particularité de fournir une analyse fine de ce qui sous-tend le fonctionnement du modèle du Drapeau Français (Davidson *et al.* 2003) ; le langage du signal et du morphogène propre à Wolpert peut donc se traduire dans le langage de la théorie des graphes, à partir du moment où le niveau d'analyse est en quelque sorte redescendu de la cellule vers les composants élémentaires du génome.

De la biologie moléculaire à la biologie du développement et la biologie cellulaire, la famille des modèles de réseau traverse en quelque sorte les

divisions de la biologie fonctionnelle ; au-delà, elle concerne même de nombreux champs d'investigation en biologie évolutive. Ainsi, les réseaux d'interaction (compétition, prédation, mutualisme..) tels que les réseaux trophiques, qu'étudie depuis plusieurs décennies l'écologie des communautés (Pimm 2002, Andrewartha and Birch 1984), partagent la nature mathématique de nombreux objets de la biologie fonctionnelle.

Notons que l'essor de ces modèles repose en partie sur les innovations technologiques qui rendent possible le recueil simultané de masses auparavant inouïes de données brutes (« big data ») - que ce soit les puces à ADN en génomique, le suivi massif d'individus par bagage et balisage en écologie, etc. - et sur leur traitement informatisé.

Dans un dernier temps nous questionnerons le rôle que peut spécifiquement jouer les objets mathématiques dans les explications.

### ***Explications mathématiques vs explications mécanistes.***

La transversalité de certains types de modèles mathématiques évoqués à l'instant laisse penser à un découplage entre l'ontologie des phénomènes (par exemple le type de « genres naturels » (*natural kinds*, Kripke) auquel on a affaire) et la mathématique des processus en jeu. Cela invite à questionner le rôle exact que pourrait jouer les mathématiques dans l'explication en biologie. On dessinait au début une opposition entre une conception philosophique « réaliste » pour laquelle les mathématiques saisiraient l'essence des phénomènes biologiques (par exemple Rosen) et une conception instrumentaliste pour laquelle ce sont simplement des outils à représenter les phénomènes biologiques. Ici il s'agit de se demander si, sans souscrire au réalisme extrême de Rosen, les mathématiques ne joueraient pas en eux-mêmes un rôle explicatif et non simplement représentatif. Ainsi, dès qu'il s'agit de graphes, certaines propriétés des graphes tels que la structure en petit monde (*Small world*, Watts et Strogatz 1998) ou en réseau invariant d'échelle impliquent des propriétés de stabilité ou de résilience que vont présenter tous les systèmes instanciant ces réseaux, indépendamment de la nature des processus mis en jeu par les entités du réseau (Huneman 2010). Dans ce cas, il semble que les propriétés mathématiques semblent de plein droit explicatives par elles-mêmes.

En philosophie des sciences, insister sur le rôle explicatif et non représentatif des mathématiques a plusieurs conséquences. D'une part, on conteste l'idée (de plus en plus fréquente parmi les philosophes de la biologie) selon laquelle expliquer un phénomène serait révéler ses mécanismes (Craver 2007, Craver et Darden 2013). Plusieurs mécanismes très différents peuvent en effet instancier des structures mathématiques identiques, et dès lors, ils donnent lieu à la même explication, laquelle apparaît indifférente envers les mécanismes. Ainsi, dans l'exemple canonique donné par Alan Baker (2009), la construction par les abeilles de cellules hexagonales régulières pour stocker leur miel est dû au fait

que l'hexagone régulier est mathématiquement prouvé comme étant la manière optimale de stocker une substance dans un espace découpé en plusieurs cellules identiques. Quels que soient les mécanismes de cognition et d'architecture des abeilles, cette propriété des hexagones – combinée au fait que l'on suppose que la sélection naturelle a pu agir sur le comportement des abeilles pour produire une structure optimale de stockage – aurait été adoptée par les abeilles. A côté de l'explication mécanistique, il faut donc reconnaître des explications mathématiques qui font abstraction des mécanismes. Dans les premières, les mathématiques ont un rôle représentatif, dans les secondes il est surtout explicatif (voir généralement Braillard et Malaterre 2014 sur ce point).

D'autre part, les philosophes discutent depuis Quine et Putnam de la thèse de « l'indispensabilité des mathématiques », selon laquelle les objets mathématiques sont nécessaires à notre conception scientifique du monde et donc jouent un rôle analogue à celui d'« entités théoriques » telles que les électrons. Si d'autres philosophes – le plus notable étant Hartry Field avec *Science without numbers* – ont contesté l'argument d'indispensabilité, reste que le cas de la biologie semble ici exemplaire. La physique est elle-même si mathématisée que les objets du physicien sont presque définis par les mathématiques (pensons au spin de l'électron, cette quantité cruciale en physique quantique quoiqu'impossible à se représenter matériellement sans mathématiques, à la différence de la masse et du volume), si bien qu'il semble difficile de conclure de l'examen de la physique contemporaine quoi que ce soit concernant la thèse de l'indispensabilité. A rebours, l'examen de la biologie permettrait de conclure, car les objets biologiques semblent à première vue ne pas être constitués par les mathématiques (l'existence d'un « chemin biochimique » semble intuitivement plus proche des concepts de la vie ordinaire que le spin de l'électron). S'il existe effectivement des explications mathématiques en biologie – que ce soit l'écologie, l'évolution ou la génomique – alors on aurait un argument fort pour affirmer au moins un type d'indispensabilité des mathématiques (sans peut-être aller jusqu'à l'espèce de platonisme auquel pensait Quine, qui repose aussi sur ses propres thèses concernant l'engagement ontologique des théories).

Ainsi, à l'inverse du sentiment de nombreux penseurs et biologistes qui nous précédèrent, l'évolution des sciences de la vie laisse voir une connexion beaucoup plus serrée entre biologie et mathématique, grâce à laquelle le rôle des mathématiques dans la compréhension scientifique en général pourrait être mieux circonscrit.

## **Conclusion.**

Comme toutes les sciences, la biologie aujourd'hui élabore des modèles mathématiques des phénomènes qu'elle étudie – que ce soit pour saisir les patterns que les données nous permettent d'appréhender, ou bien pour saisir les processus qui rendraient raison de ces patterns. La méfiance antique des biologistes et des philosophes envers les mathématiques pour ce qui est de la compréhension de la vie est aujourd'hui surmontée, aussi bien dans le champ de la biologie évolutive que dans celui de la biologie fonctionnelle, pour reprendre

la distinction classique. De même qu'à l'époque de Leibniz et Newton l'émergence du calcul infinitésimal a permis de rendre compte du mouvement, et de réconcilier mathématiques et mouvement, contre la puissante vision aristotélicienne, de même des théories mathématiques telles que celles des statistiques et des probabilités élaborées au vingtième siècle, la théorie des graphes ou encore des modèles d'équations différentielles précis tels que la réaction-diffusion ou les cycles de Lotka-Volterra ont eu raison de la résistance des biologistes envers les mathématiques. Certes, à l'inverse de la physique ou de certaines de ses branches, la biologie n'a pas de cadre mathématique général, et peut-être est-elle davantage encline à mobiliser des outils mathématiques hétérogènes sans garantie initiale d'une cohérence finale. Mais inversement, la transversalité de certains outils mathématiques en biologie et au-delà invite à repenser le rôle même des mathématiques dans l'explication scientifique. Si nous sommes loin du rêve d'une biologie mathématique de Rashevsky, reste que les mathématiques sont peut-être davantage qu'une boîte à outils commode pour résoudre certains problèmes que pose le monde du vivant.

## **Bibliographie**

- Alon U. (2007), "Network motifs: Theory and experimental approaches", *Nature Reviews Genetics*, 8: 450-461.
- Andrewartha J., Birch C. (1984), *The Ecological Web*. Chicago: University of Chicago Press.
- Baker A. (2009), "Mathematical Explanation in Science", *Brit. J. Phil. Sci.* 60: 611–633.
- Barton N., Coe J. (2009), "On the application of statistical physics to evolutionary biology", *Journal of theoretical biology*, 21, 259(2):317-24
- Bergson H. (1907) *L'évolution créatrice*. Paris: Aubier.
- Bernard C. (1865) *Introduction à l'étude de la médecine expérimentale*. Paris : Baillière.
- Braillard P. A., Malaterre, C. (2014), *Explanation in Biology : An Enquiry into the Diversity of Explanatory Patterns in the Life Sciences*. Dordrecht: Springer.
- Burnham, K. P., Anderson, D. R. (2002) *Model Selection and Multimodel Inference: A Practical-Theoretic Approach*. Dordrecht: Springer.
- Colizza, V., R. Pastor-Satorras, Vespignani A (2007) Reaction-diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics* 3:276-282.
- Craver C. (2007), *Explaining the brain*. New-York: Oxford University Press.
- Craver C., Darden L. (2013) *In search of mechanisms*. Chicago: University of Chicago Press
- D'Arcy Thompson W. (1917). *On growth and form*. Cambridge: Cambridge University Press.
- Danchin É., Charmantier A., Champagne F. A., Mesoudi A., Pujol B., Blanchet S. (2011), "Beyond DNA : integrating inclusive inheritance into an extended theory of evolution", *Nature Reviews Genetics*, 12(7), 475–486.

- Darwin C. (1859), *On the origin of species by means of natural selection*. Murray: Londres.
- Davidson E, McClay D., Hood L. (2003), "Regulatory gene networks and the properties of the developmental process." *Proc Natl Acad Sci USA* 100:1475–1480
- Davidson E.H. (1986), *Gene activity in early development*. Academic Press, Orlando.
- de la Fuente A. (2009), "What are gene regulatory networks?", In: Das S, Caragea D, Welch S, Hsu WH (eds) *Handbook of research on computational methodologies in gene regulatory networks*. Hershey, PA: Medical Information Science Reference, pp. 1–27.
- Dutreuil S. (2014), What good are abstract and what-if models? Lessons from the gaïa hypothesis. *History and philosophy of the Life Sciences*. doi: 10.1007/s40656-014-0003-4
- Falconer D. S. (1961), *Introduction to Quantitative Genetics*. Dordrecht,
- Field H. (1980), *Science Without Numbers*. Oxford: Blackwell.
- Fisher R. A. (1930), *The genetical theory of natural selection*. Clarendon, Oxford.
- Fisher R. A. (1935), *The Design of Experiments*. Oliver and Boyd. Edinburgh.
- Gaucherel, C., Munoz, F., Bérard, S. (2011). Equation against algorithm: Differences and choosing between them? *Acta Biotheoretica*, 59(1), 67–79.
- Gayon J. (1992), *Darwin et l'après-Darwin - Une histoire de l'hypothèse de sélection naturelle*. Paris: Kimé.
- Giere R. N. (1988), *Explaining science : A cognitive approach*. University of Chicago Press.
- Gillespie J. (2004) *Population genetics*, New-York, Oxford UP.
- Gotelli N.J., Graves G. (1996), *Null Models in Ecology*. Smithsonian Institution Press. Washington, DC.
- Gould S. J., Raup D. M., Sepkoski J. J., Schopf T. J., Simberloff D. S. (1977), "The shape of evolution : a comparison of real and random clades." *Paleobiology*, 3, 23–40.
- Grafen, A. (2007) "The formal Darwinism project: a mid-term report." *Journal of Evolutionary Biology*, 20: 1243–1254.
- Griffiths P. E. (2001), "Genetic information: A metaphor in search of a theory." *Philosophy of Science*, 68: 394–412.
- Heams T. (2009a), "Hérédité", in Heams, T., Huneman, P., Lecointre, G., Silberstein, M. (2009). *Les mondes darwiniens. L'évolution de l'évolution*. Syllepse, Paris.
- Heams T. (2009b), "Variation", in Heams, T., Huneman, P., Lecointre, G., Silberstein, M. (2009). *Les mondes darwiniens. L'évolution de l'évolution*. Syllepse, Paris.
- Hegel G.W.F. (1816/1994) *Science de la logique*, trad. Labarrière PJ, Gwendolyn Jarczyk G, vol. III, Paris: Aubier.
- Huneman P. (2009), "Sélection", *Les mondes darwiniens*, Heams, Huneman, Lecointre, Silberstein (eds.), Paris: Syllepses, pp.47-86.
- Huneman P. (2013), "Formal Darwinism and organisms in evolutionary biology: answering some challenges." *Biology and Philosophy*, 2014 (29) :271-279.

- Huneman P. (2014), "Mapping an expanding territory: computer simulations in evolutionary biology." *History and philosophy of life sciences*, 34 (2), 10.1007/s40656-014-0005-2
- Huneman P. (2010), "Topological explanations and robustness in biological sciences." *Synthese*, 177(2): 213–245.
- Kerszberg M., Wolpert L. (2007), "Specifying positional information in the embryo: looking beyond morphogens", *Cell* 130:205–210
- Krebs J, Davis N. (1997), *Behavioural Ecology: An Evolutionary Approach*. New-York: Oxford University Press.
- Lande R., Arnold S. J. (1983), "The measurement of selection on correlated characters", *Evolution*, 37(6), 1210–1226.
- Levins R. (1966), "The strategy of model building in population biology", *American Scientist*, 54(4): 421–431.
- Lewontin R. C. (1970), "The units of selection", *Annual Review of Ecology and Systematics*, 1, 1–18.
- Martens J. (2014), "L'analogie de l'agent économique", In Dutreuil S., Huneman P., Silberstein M., Varenne F. (eds.) *Modéliser et simuler II*, Paris: Materiologiques (sous presse)
- Maynard-Smith J. (1982), *Evolution and the Theory of Games*. Cambridge university press.
- Maynard Smith, J., R. Burian, S. Kauffman, P. Alberch, J. Campbell, B. Goodwin, R. Lande, D. Raup, Wolpert L. (1985). Developmental constraints and evolution. *Quarterly Review of Biology* 60: 265-287.
- Mayr E. (1961), "Cause and effect in biology." *Science*, 134(3489), 1501–1506.
- Mayr E. (1976), "Typological versus population thinking". In *Evolution and the diversity of life. Selected essays*. Harvard: Harvard University Press.
- McShea D.W. (1996), "Perspective : Metazoan complexity and evolution : is there a trend?", *Evolution*, 50(2): 477–492.
- Merlin F. (2013), *Mutations et aléas : le hasard dans la théorie de l'évolution*. Paris: Hermann.
- Metz JAJ (2008), "Fitness", In Jørgensen SE, Fath B(eds.), *Evolutionary Ecology*, 2. pp.1599-1612. Elsevier.
- Morgan M. S., Morrison M. (1999), *Models as mediators*. Cambridge: Cambridge University Press.
- Nowak M. (2006) *Evolutionary Dynamics: Exploring the Equations of Life*. Harvard : Belknap Press
- Okasha S. (2006), *Evolution and the Levels of Selection*. Oxford University Press, USA.
- Pimm S. (2002), *Food webs*. Chicago: University of Chicago Press (2<sup>nd</sup> ed.)
- Price G. (1970), "Selection and covariance." *Nature* 227: 520–521.
- Rashevsky N. (1938), *Mathematical biophysics. Physicomathematical foundations of biology*. University of Chicago Press.
- Rosen R. (1991) *Life Itself: A Comprehensive Inquiry into the Nature, Origin, and Fabrication of Life*. New-York : Columbia University Press.
- Rosenberg A., Bouchard F (2010) "Fitness", *Stanford Encyclopedia of E. N.Zalta* (ed.) <<http://plato.stanford.edu/archives/fall2010/entries/fitness/>>
- Salmon W. (1989), *Four Decades of Scientific Explanation*. Pittsburgh : University of Pittsburgh Press.



- Sella G., Hirsh A.E. (2005), "The application of statistical physics to evolutionary biology", *PNAS* 102: 9541–954
- Short M., Brantingham J., Bertozzi A., Titad G. (2010), "Dissipation and displacement of hotspots in reaction-diffusion models of crime", *PNAS*, 107(9), 3961-3965.
- Sober E. (2002), "What is the problem of simplicity?" In A. Zellner, H. A. Keuzenkamp, McAleer M. (Eds.), *Simplicity, inference and modelling: Keeping it sophisticatedly simple* (pp. 13–31). Cambridge: Cambridge University Press.
- Strogatz S. (2001), "Exploring complex networks." *Nature* 410, 268-276.
- Tompkins, N., Li, N., Girabawe, C., Heymann, M., Ermentrout, G. B., Epstein, I. R., Fraden, S. (2014), "Testing turing theory of morphogenesis in chemical cells." *Proceedings of the National Academy of Sciences*, 111(12), 4397–4402.
- Turing A. (1952), "The chemical basis of morphogenesis." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 237(641):37–72.
- Turner D. (2011), *Paleontology: A Philosophical Introduction*. Cambridge : Cambridge University Press.
- Van Frassen B. (1980) *The scientific image*. New-York : Oxford University Press.
- Vorms M. (2011), *Qu'est-ce qu'une théorie scientifique?* Vuibert.
- Walsh D. (1996), "Fitness and function." *Philosophy of science*. 47 (4):553-574.
- Watson A., Lovelock, J. (1983), "Biological homeostasis of the global environment : the parable of Daisyworld." *Tellus B*, 35(4): 284–289.
- Watts D. J., Strogatz S. H. (1998), "Collective dynamics of small-world networks", *nature*, 393(6684), 440–442.
- Werndl, C. (2013). Climate models, confirmation and calibration. *British Journal for the Philosophy of Science*, 64: 609–635.
- Wolpert L. (1969), "Positional information and the spatial pattern of cellular differentiation", *Journal of theoretical biology*, 25(1), 1–47.
- Wright S. (1932), "The roles of mutation, inbreeding, crossbreeding and selection in evolution", *Proceedings of the sixth annual congress of genetics* 1: 356-366